



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

What is speech rhythm?

Citation for published version:

Turk, A & Shattuck-Hufnagel, S 2013, 'What is speech rhythm? A commentary inspired by Arvaniti & Rodriquez, Krivokapi, and Goswami & Leong.', *Laboratory Phonology*, vol. 4, no. 1, pp. 93-118.
<https://doi.org/10.1515/lp-2013-0005>

Digital Object Identifier (DOI):

[10.1515/lp-2013-0005](https://doi.org/10.1515/lp-2013-0005)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Laboratory Phonology

Publisher Rights Statement:

With permission. © Walter de Gruyter. Turk, A., & Shattuck-Hufnagel, S. (2013). What is speech rhythm?: A commentary inspired by Arvaniti & Rodriquez, Krivokapi, and Goswami & Leong. *Laboratory Phonology*, 4(1), 93-118doi: 10.1515/lp-2013-0005

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Alice Turk and Stefanie Shattuck-Hufnagel

What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong

Alice Turk: The University of Edinburgh, Edinburgh, UK. E-mail: turk@ling.ed.ac.uk

Stefanie Shattuck-Hufnagel: Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology

1 Introduction

Speech rhythm has been the focus of a great deal of research over the past 80 years. The papers in this section provide a good snapshot of the current state of this research, because they illustrate the range of topics that the term ‘rhythm’ covers, as well as the range of its possible definitions. One of the issues that these papers highlight for us is the fact that the term ‘rhythm’ can mean very different things to different people. We believe that the use of this term in reference to speech, if not further qualified, can lead to misunderstandings about what claims are being made about the fundamental nature of speech. These misunderstandings are potentially serious because the term ‘rhythm’ carries with it implicit assumptions about the way speech works, and about how (if at all) it involves periodicity. As a result, readers who come to a paper on speech rhythm with a different definition in mind, or with different implicit assumptions about the role of periodicity in speech, may take away a very different impression from what the authors intend. For example, a proposal for oscillator-based perceptual mechanisms described as detecting ‘rhythmicity’ in the speech signal could easily be understood as detecting (and therefore requiring) ongoing periodicity in the speech signal. On the other hand, if the proposed mechanism involves locally tuning the periods of a perceptual oscillator so that they each time-lock to inter-event intervals in the speech signal, even though inter-event intervals might be of different durations, the claim about the temporal structure of spoken utterances is very different: periodicity in the signal is no longer required.

The authors of these three papers all suggest that there is a sense in which normal conversational speech can be considered periodic (although they have very different views of where in the speech chain this periodicity plays a role). But how persuasive is the evidence for this claim that typical communicative speech

involves periodicity? On our view, given the extensive evidence that normal conversational speech is not periodic on the surface, i.e., that no constituent recurs at regular temporal intervals, it is still very much an open question whether or not speech is (1) controlled using periodic control structures, and/or (2) perceived as periodic. Although the latter claim is widely assumed, it has been subjected to remarkably little empirical testing, and so must be taken as a hypothesis rather than an established observation. In Section 2 of this paper we discuss the range of definitions of rhythm and which one(s) might be appropriate for speech.

A second question raised by these three papers concerns rhythm typology. It is certainly the case that some people have strong intuitions about what might be called the *global rhythmic* (or *timing*) profiles of different languages. However, what contributes to these intuitions is less clear, as is their generality, i.e., whether they are shared by all listeners, and for all utterances of a language. The papers by Arvaniti and Rodriquez and Krivokapić address this issue, and show that it is far from straightforward to empirically substantiate earlier-reported intuitions about rhythm type (Lloyd James 1935, 1940, cited in Dellwo 2010; Pike 1945; Abercrombie 1967) that are often taken for granted. We consider this question more thoroughly in Section 3 below.

Finally, in Section 4, we propose that a fruitful initial approach to studying ‘speech rhythm’ involves studying speech timing more generally: its control structures and processes, its perception, and the systematic relationship between phonology (both segmental and prosodic) and surface timing. This approach incorporates the wide variety of factors influencing speech timing into the ongoing search for rhythm in speech and rhythm classes for languages. It takes account of the many factors that may influence a listener’s sense of the *global rhythmic profile* of a language (which we will argue might be termed the *global timing profile*), as well as a similarly wide (but not identical) variety of factors that may influence the *timing pattern of a specific utterance* in that language. As will become clear below, even though timing is not the only aspect of spoken utterances that might contribute to a listener’s sense of their rhythm, we believe that an understanding of how speech timing works is a necessary prerequisite to understanding rhythm in all its possible meanings.

2 What definition of ‘rhythm’ is appropriate for speech, and what role does periodicity play in speech processing?

As questions about speech rhythm have become a topic of re-invigorated quantitative investigation in the speech research community, the general assumption has been that speech is rhythmic, and the focus has been on how speech rhythmicity might be manifested. However, if one steps back from this approach to ask the prior question of whether or not speech **is** rhythmic, it turns out that the answer is very hard to specify, because there is a wide range of meanings for this term. So, our first issue is, what are the possible definitions of ‘rhythm’? It appears that there are three major approaches to understanding the meaning of this term as applied to speech. In all of these approaches, ‘rhythm’ refers to some aspect of timing, but the definitions differ as to whether rhythm (1) includes some aspect of periodicity in timing, (2) refers to abstract structurings of time, based on serial ordering, that reflect grouping and prominence structure, and/or (3) refers to systematic surface timing patterns determined by grouping and prominence structure as well as other factors. We address each of these types of definition in turn.

2.1 Definition set 1: Rhythm based on periodicity

This set of definitions invokes periodicity either in motor or perceptual processing, but the definitions differ in the extent to which this periodicity must be measurable on the surface.

2.1.1 Definition 1a: Surface periodicity

One subset of periodicity-based definitions that emphasize the role of linguistic planning units and processes (Abercrombie 1967) proposes that surface (near)-periodicity in the signal is imposed by periodicity-based control structures, i.e., that some linguistic constituent in the speech signal is planned to occur with temporal periodicity. Any variations in surface inter-event or constituent durations may be expected to reflect either (a) noise at the motor implementation level (e.g., as discussed in Wing and Kristofferson 1973) that occurs in spite of planned isochrony, or (b) ‘silent’ beats, where surface events reflecting the isochronous pulse are omitted (Abercrombie 1967).

2.1.2 Definition 1b: Periodicity-based motor control

A second subset of periodicity-based definitions emphasizes the role of periodicity-based motor control mechanisms (Classe 1939; O'Dell and Nieminen 1999; Barbosa 2007; Saltzman et al. 2008; Kohler 2009), in contrast to planned surface periodicity as discussed above. It hypothesizes that periodicity-based motor control is evidenced by a tendency towards isochrony of some constituent in the surface signal. Such evidence would include polysegmental and polysyllabic shortening, where a larger number of subconstituents in a constituent (e.g., segments or syllables in an inter-stress interval or word) results in temporal compression of these segments or syllables. The resulting constituent duration is shorter than would be expected based on the number of subconstituent segments and/or syllables. This surface shortening pattern might arise from a 'rhythmic tendency' which "has to contend with other factors which obscure its effects" (Classe 1939: 87), or from isochrony in production control in the form of the interaction of coupled planning oscillators at multiple levels of prosodic constituency (syllable, foot, and phrase; see O'Dell and Nieminen 1999; Barbosa 2007; Saltzman et al. 2008), whose activity is presumably uncovered by the listener's perceptual processes.

2.1.3 Definition 1c: Periodicity-based perception

A third subset of periodicity-based definitions emphasizes the role of periodicity in perception. On these views, periodicity-based perceptual processing unfolds with little requirement for surface periodicity in the signal, or for periodicity in motor control. For some researchers (e.g., Lehiste 1977¹), periodicity-based perception involves the impression of periodicity: even a signal that is largely aperiodic is heard as isochronous.

More recent work has emphasized the role of neural oscillators in speech perception (Ghitza 2011; Giraud and Poeppel 2012; Goswami and Leong, this volume; see also Large 2010; Loehr, Large, and Palmer 2011 for music perception). This work has been inspired by observations from neuroscience that the brain produces oscillations in its default state. In some of these models (e.g., Ghitza 2011, based on Ahissar et al. 2001; Nourski et al. 2009), oscillator periods remain time-locked to the inter-event intervals that they track, and therefore change according

¹ Lehiste (1977) leaves open the possibility that there is also a periodic component in motor control.

to the inter-event durations in the signal. Although it is not clear exactly how temporal percepts are derived from neural oscillator periods in these models, they do not necessarily imply periodic impressions. As Goswami and Leong (this volume) propose, they do, however, help explain a possible link between language disorders and behavior on periodic tasks, since normal, aperiodic speech perception and periodic tasks are proposed to both make use of perceptual oscillators.

2.2 Definition 2: Abstract phonological rhythm

A second definition of speech rhythm is based not on periodicity, or even on any aspect of surface timing per se, but instead on abstract structuring of points in time, corresponding to serially ordered linguistic elements with relative salience, that reflect prominence and grouping structure (Kozhevnikov and Chistovich 1965; Liberman 1975; Liberman and Prince 1977; Giegerich 1980; Selkirk 1984; Cutler 1994; Murty, Otake, and Cutler 2007; Arvaniti 2009). By this definition, timing is but one of many correlates of abstract prosodic prominence and grouping structure (Arvaniti 2009). One example of this view, proposed by Liberman (1975) and Liberman and Prince (1977), posits that the morpho-syntactic structure of an utterance undergirds the formation of a metrical grid of timing slots, where the relative salience of elements associated with each timing slot is indicated by the heights of the columns in the grid. The metrical grid reflects the grouping and prominence structure of the utterance in abstract ways. For example, prominence structure is reflected in the height of the bars on the grid, but the phonetic correlates of this structure are not specified. Furthermore, the grid does not directly represent grouping structure, but instead indirectly reflects it in two ways: (1) relative prominence is influenced by constituent structure, e.g., the strongest prominence (nuclear stress) is stipulated to associate with rightmost phrasal nodes (Liberman 1975: 196), and (2) the grid can also include extra timing slots at phrasal boundaries, where phonetic correlates can include either final lengthening, pause, or both (Liberman 1975; Selkirk 1984). The fact that grouping structure is reflected in the grid means that Liberman's view of rhythm cannot be equated with prominence structure alone, but instead more broadly includes reflexes of prominence and grouping structure together.

Although this abstract phonological definition of speech rhythm does not make specific predictions about the surface timing or other acoustic correlates of prominence and grouping structure, it is nevertheless compatible with definitions of rhythm that focus on systematic patterns in surface timing, because prosodic grouping and prominence structure are known to affect the surface timing

characteristics of utterances in predictable ways. In the surface-timing-based definitions of rhythm that we discuss next, rhythm can be seen as the concrete phonetic specifications for the timing pattern of an utterance that result from abstract timing specifications (e.g., from a metrical grid) in combination with additional factors affecting timing, e.g., segments and overall rate of speech.

2.3 Definition 3: Rhythm as surface timing patterns

The third definition focuses on systematicity in surface timing. On this view, speech is characterized by timing patterns that, while not necessarily isochronous, are the result of systematic relationships between factors such as segmental identity and context (both segmental and prosodic) on the one hand, and surface duration patterns on the other (Dasher and Bolinger 1982; Dauer 1983; Ramus, Nespor, and Mehler 1999; Low, Grabe, and Nolan 2000). Some researchers focus on a particular subset of these factors. Dauer (1983), for example, in a seminal paper, selected vowel reduction, lexical stress, and syllable-related phonotactics as having a particularly strong influence on language-specific surface timing patterns. Other research has shown that other factors, e.g., overall rate of speech, segment identity, grouping, and phrasal prominence, can have large effects on timing patterns as well (e.g., Lehiste 1970; Wightman, Shattuck-Hufnagel, Ostendorf, and Price 1992; Turk and White 1999; Turk and Shattuck-Hufnagel 2007; and many other papers cited in Fletcher 2010), although cross-linguistic differences in these effects are less-well documented.

Ramus et al. (1999), following Mehler et al. (1996), make the strong claim that rhythm is a surface phonetic property only, reflected in the durations and intensities of vocalic and consonantal intervals. On this view, rhythm perception doesn't require knowledge of the abstract phonological structures to which it relates: even infants can compare the rhythms of languages that they hear. Other researchers assume that rhythm includes both surface phonetic timing patterns and abstract prosodic structure. For example, Patel (2008: 96) refers to rhythm as "systematic patterning of sound in terms of timing, accent and grouping", which combines abstract prosodic structuring of time with systematicity in surface timing. And Arvaniti's (2009) views are consistent with a different hybrid definition, which combines abstract prosodic structuring of time with periodicity in perception.

In considering which of these definitions of rhythm (or views of timing) is appropriate for speech, an important question is whether speech is periodic or not, and if so, in what way. That is, does normal, conversational speech involve periodicity? There is no doubt that speech is 'rhythmic' in the sense that it is seri-

ally ordered and that the serially ordered elements of speech are structured into hierarchical grouping and prominence structures involving stronger and weaker elements (see Fletcher [2010] and Shattuck-Hufnagel and Turk [1996] for reviews), as in definition 2, and moreover that there are systematic relationships between grammatical and performance factors on the one hand, and surface duration patterns on the other (e.g., Lehiste 1970; Nooteboom 1972; Klatt 1976; van Santen 1992; van Santen and Shih 2000; Jurafsky, Bell, Gregory, and Raymond 2001; Bybee and Hopper 2001; Bell et al. 2003; Aylett and Turk 2004; and many references mentioned in Shattuck-Hufnagel and Turk 1996 and Fletcher 2010), as in the systematic surface timing view (definition 3). But is speech ‘rhythmic’ in the sense of being governed by periodicity-based production or perception mechanisms, as in the definitions in set 1?

2.4 The question of periodicity

2.4.1 Lack of evidence for surface periodicity in normal conversational speech

Given the overwhelming evidence for the influence of grammatical and performance factors on duration, it is perhaps not surprising that empirical tests of surface isochrony for elements such as syllables and inter-stress intervals in different languages have failed to find evidence for it. For example, inter-stress intervals in English and Dutch were found to be far from isochronous (Uldall 1971; Dauer 1983; den Os 1988, and many others cited in Fletcher 2010). Instead, the inter-stress interval duration increases with the number of syllables it contains (O’Connor 1968; Lea 1974; Thompson 1980, cited in Scott, Isard, and de Boysson-Bardies 1985; as well as many other references cited in Dauer 1983 and Fletcher 2010). Moreover, careful measures of utterances in languages characterized as ‘syllable-timed’ showed syllable duration variability that was just as large as in stress-timed languages, and inter-stress interval duration variability that was even larger in stress-timed languages than in syllable-timed languages (Roach 1982; see also Delattre 1966). This means that periodicity-based definition 1a (surface isochrony) is untenable, at least for normal conversational speech. However, as we discuss below, it is possible that surface periodicity can be selected as a production goal by the speaker for particular reasons, in certain styles of speech, such as poetic or ‘rhythmicized’ speech.

2.4.2 Periodicity in motor control?

Periodicity in motor control (1b) is attractive because some aspects of motor activity, e.g., chewing and locomotion, appear to be periodic, suggesting that periodicity may be basic to human motor activity. Although speech is not normally periodic on the surface, some phenomena are interpretable if periodicity is viewed as part of speech planning or motor control processes. For example, Fant and Kruckenberg's (1989) finding that final lengthening + pause durations together appear to be an integer multiple of the average inter-stress interval duration in Swedish is explained by a planning process that includes an internal pulse rate synchronized to a "local average of stress rate" (60). As Krivokapić (this volume) discusses, the Articulatory Phonology/Task Dynamic model of speech motor control (Saltzman et al. 2008, following O'Dell and Nieminen 1999 and Barbosa 2007) includes periodic control mechanisms. Even though this model has periodic syllable, foot, and phrase control structures, the model does not require actual surface periodicity: the surface durations that result from these control structures need have only a tendency toward isochrony. Polysyllabic and polysegmental shortening (shortening of segments or syllables when more are included in an interval) are cited as the main evidence for tendencies towards isochrony, and therefore for periodicity in models of speech production (Classe 1939; Pike 1945; Lehiste 1972). For example, English is often claimed to have polysyllabic shortening within inter-stress intervals (Williams and Hiller 1994; Kim and Cole 2005; Kim 2006), supporting the proposal of periodic control at an inter-stress interval level. However, many of these studies did not examine the alternative possibility that polysyllabic shortening applies within word-based units, such as content + function word units, in preference to inter-stress intervals (but see Kim [2006], which tested the possibility that it occurs within words, but did not test the possibility that it occurs within content + function word units, rather than within inter-stress intervals). One preliminary study which did control for different possible domains of polysyllabic shortening found more support for word-based prosodic units than for inter-stress intervals, even in poetic contexts, where inter-stress interval periodicity involving word fragments would be most likely to surface (Shattuck-Hufnagel and Turk 2011). This result puts polysyllabic shortening in the domain of grammatical-structure-based, rather than periodicity-related, timing adjustment. Furthermore, White and Turk (2010) and Turk and Shattuck-Hufnagel (2000) found that polysyllabic shortening does not occur in all contexts, preferentially targeting phrasally stressed words, where it is also greatest in magnitude. These results support the idea that rather than being a reflection of a periodic control mechanism tending toward surface isochrony, polysyllabic shortening may instead be one of a set of mechanisms that speakers

use to signal, on a subset of words only, the locations of word-based prosodic constituent boundaries in an utterance (Turk 2012).

These findings suggest that, despite the impressive modeling power of the Articulatory Phonology/Task Dynamics approach, there are reasons to be uncertain whether periodicity is a major factor in speech motor control in typical speaking circumstances. However, even if this doubt turns out to be well founded and periodicity plays no role in normal conversational speech, it is still possible that periodic control structures (or surface periodic planning goals) are invoked for certain kinds of speech production that might be called rhythmicized or periodized speech, such as singing or the reciting of limericks, or for stylistic purposes during typical communicative speech. The possibility that the timing pattern for an utterance could be computed by different mechanisms, depending on the task requirements (e.g., periodic vs. non-periodic requirements) is deserving of future experimental treatment.

2.4.3 Periodicity in perception?

We turn now to a different aspect of the periodicity question: the possibility that the impression of periodicity comes from the perceptual system, as noted above in the third periodicity-based definition of speech rhythm (1c). Given the lack of evidence for surface periodicity, periodic perception seemed to many investigators to be a possible explanation for the widespread sense of rhythmic regularity that many listeners share, and was explicitly proposed by Lehiste (1977) and others. If this view of perceptual periodicity is correct, then it is important to ask why and how this perceptual mechanism imposes regularity on a surface-aperiodic signal. A potential explanation is suggested by a possible extension of the Articulatory Phonology/Task Dynamics model (discussed above and in Krivokapić, this volume) into the perceptual domain. As we have seen, this model proposes that timing in the speech signal is accomplished by a set of coupled oscillators, e.g., syllable-, foot-, and phrase-level planning oscillators, whose relative coupling strength determines the timing of these linguistic constituents. Extending this model into the perceptual domain would involve postulating that listeners uncover these oscillators and their frequencies when perceiving the speech. Such an approach has the advantage of accounting for the strong intuition felt by many listeners that spoken utterances are rhythmically regular despite their lack of surface periodicity, so that languages can differ in the type of regularity they show.

However, as discussed above, the facts about polysyllabic shortening challenge the motivation for this type of oscillator-based model in the production domain. And even leaving this problem aside, other facts about speech timing have

led to developments within the Articulatory Phonology model that suggest that the perceptual recovery of the syllable-, foot-, and phrase-level oscillations would be anything but straightforward. In particular, constituent-initial and -final lengthening and phrasal-stress-related lengthening have required the postulation of timing modulation mechanisms (Pi and μ_r lengthening gestures) at boundaries and prominences (Byrd and Saltzman 2003; Saltzman et al. 2008). The perceptual system would need a way to look beyond these temporal perturbations to recover the weighted, interacting frequencies of the syllable-, foot-, and phrase-level planning oscillators.

Even if periodicity is not part of the motor control system for typical conversational speech, it is possible that the perceptual system imposes regularity on the aperiodic speech signal. This hypothesis is largely unexplored. Lehiste (1977) tested how well listeners could compare the durations of a sequence of inter-stress intervals, and found that listeners required large differences in duration in order to reliably judge which of four consecutive intervals was the shortest or longest. However, judging which of a sequence of intervals is longest is a very different task from judging whether a sequence sounds isochronous, so this result may not bear on the question of whether listeners regularize irregularities in successive temporal intervals. Donovan and Darwin (1979); Darwin and Donovan (1980), cited in Scott et al. (1985); and Scott et al. (1985) tested how accurately listeners could tap to the stressed syllables in spoken utterances, and found that they tended to regularize the irregular intervals between taps. This might suggest a perceptual regularization mechanism, but little is known about the conditions under which such regularization occurs. Scott et al. (1985) suggest that regularization is a response bias that comes into play when stimuli are complex and the perception task is difficult.

It appears that, if perceptual regularization does occur, the process is somewhat complex. For example, regularization of durations at a very early stage in perception would eliminate some of the timing cues for segment contrasts and word boundary locations. This would present a challenge to the growing body of evidence suggesting that segment, syllable, and context durations play a role in segment and word identification (see Miller 1981 for a review; Newman and Sawusch 1996; Shatzman and McQueen 2006; Casini, Burle, and Nguyen 2009; Reinisch, Jesse, and McQueen 2011), and in word segmentation (Salverda, Dahan, and McQueen 2003; Shatzman and McQueen 2006). If perceptual regularization of timing occurs, it must be in a way that does not interfere with these processes. In this regard, it may be important to consider the difference in scale between the very short durations relevant to segmental processing and the considerably longer ones relevant for inter-stress intervals, which may invoke separate processing mechanisms.

One possibility is that durations of various intervals in speech are processed without periodic regularization at an early stage of processing, and that this un-regularized timing information is subsequently passed to other processing mechanisms, such as those that support (1) grammatical inference, i.e., identification of contrastive linguistic elements, (2) rate/style of speech processing, and (3) periodicity processing. In this view, perceptual regularization would be part of (3) periodicity processing. This view provides a possible account of the intuition that speech is rhythmic in the sense of perceptual isochrony of some constituent on the surface, even though listeners are able to use non-regularized durations for grammatical purposes. That is, speakers of a language may understand how timing is systematically governed by segments, grouping and prominence hierarchies, and performance factors. As a result, at least in typical communicative circumstances, they use the timing information in an utterance to determine its segmental, prosodic, and performance structure, and don't pay much attention to its relationship to periodicity. In addition, in particular circumstances a listener may focus on the potential for an utterance to be analyzed in terms of periodicity, e.g., when the speaker structures an utterance so that it exhibits quasi-periodicity on the surface in terms of linguistic elements such as syllables, feet, or phrase-level pitch accents (Kohler 2009, following Classe 1939; Nolan and Jeon to appear), or when the listener is dealing with an utterance in a language for which the segmental, prosodic, and performance structures are not second nature, so that timing information is processed in a different (periodic) way. Relatedly, it is possible that the sense that listeners have that some speakers are more 'rhythmic' than others is due to the ability those speakers have to signal the grouping and prominence relations of their utterances in a particularly fluent and transparent way, rather than because they more faithfully follow a rhythmic pulse pattern.

In sum, the definition of rhythm that one adopts for speech, and its implications for the role of periodicity in speech processing, are critical for understanding and interpreting experimental results. For this reason, it is important for investigators who address the question of speech rhythm to specify both of these aspects of the framework within which they are working. Although widely-adopted models of speech timing control rely on the concept of periodicity in motor control (as in 1b), much of the evidence viewed as support for these models has plausible alternative interpretations. Thus, explicit experiments are required to distinguish between such models, in which periodic mechanisms are used to control the timing of speech movements, vs. competing models in which periodicity plays no role in normal conversational speech. Finally, models of perception based on periodic oscillators can explain some perceptual regularization phenomena, but do not explain the widely documented use of timing cues for gram-

mathematical purposes (see also Cummins 2012). In short, much of the critical work in this area remains to be done, and will benefit from explicit statement and exploration of assumptions. In the next section we turn to the question of where this leaves the issue of rhythm typology.

3 Rhythm typology

Much of the interest in rhythm in recent years stems from early claims that languages of the world could be classified in terms of ‘rhythm classes’ (Lloyd James 1935, 1940, cited in Dellwo 2010; Pike 1945; Abercrombie 1967; henceforth L-JPA). Indeed, it can be said that many of the definitions of rhythm reviewed in Section 2 resulted from attempts to distinguish among languages that were thought to belong to different rhythm classes.

The claims of the L-JPA ‘rhythm hypothesis’ were three-fold:

1. that languages sound different in part because of their rhythm distinctions,
2. that these different percepts of rhythm reflect the isochrony (Lloyd James 1935, 1940; Abercrombie 1967) or near-isochrony (Pike 1945) of different linguistic units for different languages, and
3. that languages fall into two (or three) rhythm classes, depending on the units which either are isochronous or predominate in their tendency towards isochrony (Pike 1945). For example, English, Russian, Arabic, and Persian have been classified as stress-timed, predicting isochronous or near-isochronous occurrence of stressed vowels, while French, Spanish, Telugu, and Yoruba have been classified as syllable-timed. Bloch (1950), Han (1962), and Ladefoged (1975) proposed a third category, mora-timing, exemplified by Japanese.

Although there is an accumulation of evidence against surface isochrony in any language, there is a reluctance to abandon the idea that the prototypical syllable- and stress-timed languages identified by L-JPA are rhythmically different. Three proposed alternatives to the original L-JPA view preserve the notion of rhythmic differences, despite the lack of surface isochrony. These reflect various of the definitions presented in Section 2.

1. Rhythmicity determined by production control processes (periodicity in motor control, definition 1b above): In this view, rhythmic types reflect underlying periodic control processes which result in tendencies towards isochrony of different units, without reaching actual surface isochrony (Pike 1945; O'Dell and Nieminen 1999; Krivokapić, this volume). Articulatory Phonology/Task Dynamics models which rely on the interaction of oscillator-based planning elements to control timing can deal with this tendency, as we have

seen, by invoking different strengths of coupling between planning oscillator levels.

2. Rhythmicity determined by perceptual processes (periodicity in perception, definition 1c above): On this view, different rhythmic types reflect the way that the *global rhythmic* (or *timing*) *profiles* of different languages are perceived. This may involve either the perception of utterances as isochronous sequences of units of different types (as suggested by Lehiste [1977] for English inter-stress intervals) or in other ways (Arvaniti and Rodriquez, this volume).
3. Rhythmicity determined by phonological structure (definitions 2 and 3 above): On this view, languages of different rhythmic types can be distinguished by phonological properties which have measurable influences on timing in speech acoustics and/or prominence placement (Dauer 1983; Selkirk 1984; Ramus et al. 1999), but are not distinguished by the surface isochrony of different types of units. Dauer (1983) and Nespor (1990) propose that within each rhythm class, a set of phonological properties converges in its influence on surface phonetics to give the impression of syllable- vs. stress-timing, whereas Selkirk (1984) proposes that the difference between syllable- and stress-timed languages lies in the metrical grid, where beats can occur on every syllable in syllable-timed languages, but on only some syllables in stress-timed languages.

On all of these views, languages can be classified into discrete rhythm categories, which is a natural interpretation of the term ‘rhythm class’ or ‘rhythm typology’. However, faced with the complexities revealed by close durational analysis of actual utterances, a number of researchers have proposed that languages might not fall into a small number of categorically discrete rhythm classes, as proposed by L-JPA, but instead might fall at particular locations on a continuum of rhythm-related parameter values, and that these parameters might reflect a number of different dimensions rather than just one (Nespor 1990; Grabe and Low 2002).

As this brief summary shows, the overarching goal in most rhythm typology work has been to figure out why syllable-timed and stress-timed languages (as identified by L-JPA) sound different. A popular approach has been to assume that the basic difference proposed by L-JPA is correct, and to search for acoustic or structural correlates of this perceptual difference. If acoustic measures could be found to distinguish the L-JPA categories, these would provide insight into what physical characteristics in the speech signal are responsible for the perception of these ‘rhythmic’ differences. Initial results were promising. However, as more detailed studies were carried out, problems and limitations of this approach emerged, as we discuss below.

3.1 Rhythm metrics

The search for an acoustic correlate of the L-JPA intuition led to a number of proposed rhythm metrics (discussed in White and Mattys 2007; Wiget et al. 2010; Loukina et al. 2011), based on durations of different types of intervals. Some are defined acoustically (e.g., vocalic and consonantal intervals; Ramus et al. 1999; Low, Grabe, and Nolan 2000; Grabe and Low 2002; Dellwo and Wagner 2003), while others are defined in terms of phonological constituents (e.g., syllables and feet; Deterding 2001; Barry et al. 2003; Nolan and Asu 2009). Some of the metrics (in particular those involving standard deviations) are normalized for speech rate, to control for the correlation observed between speaking rate and the metrics (Barry et al. 2003; Dellwo and Wagner 2003, discussed in White and Mattys 2007). Some quantify variability in intervals over entire speech samples (Ramus 2002), while others measure the negative co-variance of adjacent intervals (Grabe and Low 2002). These metrics have been tested separately and in combination, to see how effectively they distinguish languages believed to be rhythmically different. Many of the metrics (or combinations) were initially successful, in that they separated languages traditionally classified as syllable-timed from languages traditionally classified as stress-timed. On the basis of this initial success, other as-yet-unclassified languages were analysed so that they could be situated on the continua provided by these metrics, with the idea that position on the metrics could be used to ‘classify’ these languages.

3.1.1 Problems that make rhythm metric results difficult to interpret

Although rhythmic measures research succeeded in quantifying some aspects of timing-related differences among languages, several issues make it difficult to interpret the findings in terms of rhythm types or classes.

3.1.1.1 Lack of empirical testing of the widely held belief in rhythmic differences among languages

Rhythm metric research is based on the assumption that the L-JPA impression of rhythmic differences among languages is correct, but this impression lacks empirical support. That is, quantitative tests of what causes people to call two languages (or utterances) rhythmically similar or different is lacking. Arvaniti and Rodriquez’s paper (this volume) and White, Mattys, and Wiget (2012) point out that one such perceptual test, the flat- F_0 *sasasa* test of Ramus and Mehler (1999) and Ramus, Dupoux, and Mehler (2003), fails to provide evidence for rhythm

classes. This may be because the number and type of stimulus sentences were insufficient to allow listeners to form a *global rhythmic* [or *timing*] *profile* of each language. But whatever the reason for this result, in the absence of such evidence, it is challenging to interpret results of rhythm metric classification tests. For example, imagine two languages that give exactly the same value on the standard deviation of vocalic interval duration, but where one of the languages has only vowel reduction and no boundary- or accent-related lengthening, while the other language has no vowel reduction and only boundary/accent-related lengthening. These two languages might give the same standard deviation of vocalic interval duration measure – but would one want to say they are of the same rhythmic type? Perhaps one would, if the rhythmic impression they give to a listener is the same; on the other hand, one might not, if the rhythmic impression they give is different, because, e.g., the listener can distinguish between a duration pattern created by vowel reduction and one that reflects boundary/accent-related lengthening (see Arvaniti 2009: 55 for a similar discussion). Thus, it would be important to test the degree of similarity in the rhythmic impression produced by utterances in the two languages. Devising a quantitative test of this kind might be challenging, since the choice of specific utterances in a necessarily-limited experimental test would be critical to the results. And in the end, one would want to know about the difference in underlying causation, which might require an understanding of how the various factors that influence timing in a given language do their work.

If an appropriate quantitative test could be found, it would also be important to use it to determine whether all listeners share the L-JPA intuitions, or even whether they hold for all or most native English listeners. Even if they hold for all native listeners, would they be different for native listeners of other languages? Would this classification be different if listeners were listening to languages that they knew, e.g., do French and English sound ‘rhythmically different’ to French-English bilingual listeners? It is possible that a bias toward periodic perception might be discouraged when listening to a language whose grammatical structures are known to the listener, so that the timing relationships make sense in terms of those structures, but encouraged when listening to a language whose grammatical structures are unknown. Thus, it might be useful if searches for the measure that correlates with the intuitions about rhythm type differences were preceded by (or at least accompanied by) experimental demonstrations of the generality of the intuitions about each language type.²

² In a recent set of experiments, Arvaniti (2012b) found no evidence that listeners from different native language groups differed in their rhythm judgments when grammatical information was unavailable, i.e., when stimuli were low pass filtered, or converted to re-iterant (‘flat *sasasa*’)

3.1.1.2 The nature of the typological space is unclear, leading to problems of rhythm metric interpretation

On a related point, it is unclear whether there is a continuum of rhythmic impressions, or a small set of qualitatively different impressions. That is, even if it turns out that the L-JPA impression of a rhythmic difference between English, Arabic, and Russian on the one hand and French, Spanish, and Telugu on the other is correct, it is unclear whether these languages are points on a continuum of rhythmic attributes, or are representative members of a set of distinct rhythm classes. Without knowing which of these two types of perceptual impression we are trying to find the correlate for, the typological interpretation of results for particular physical rhythm metrics is difficult if not impossible. For example, if a language is found to be intermediate between two typological categories according to a particular (set of) rhythm metric(s), it is unclear whether this is because the typological space is continuous and the language is appropriately situated in the middle of the continuum, or because the chosen rhythm metric has not successfully identified the language's membership in its appropriate rhythm class.

3.1.1.3 Need for an adequate sample of utterances from a language to determine its rhythm

It is unclear how much material is required to assess the 'rhythm of a language'. The Abercrombie claim was that typological differences related to differences in isochronous production units. This hypothesis suggests that rhythmic differences should be observable from a very small sample of utterances from any two languages belonging to different rhythmic classes. However, those who think that rhythm relates to systematicities in surface patterning due to phonological categories, structure, and their distribution (e.g., Dauer 1983) might predict that successful quantification of language rhythm would need to be based on the statistical analysis of a large corpus that is representative of the language as a whole (cf. Arvaniti 2009; Wiget et al. 2010; Loukina et al. 2011; Arvaniti 2012a; Renwick 2012). The need for an adequate sample of utterances is reinforced by the observation that utterances of individual sentences from a given language may vary strikingly in their value on a rhythm metric (Arvaniti 2009, 2012a; Wiget et al. 2010; Renwick 2012), and that the rhythm-related characteristics of a given sentence can vary from speaker to speaker according to speaking skill (Kohler 2009).

speech. This result is consistent with the rhythm typology view. However, she also found no evidence that listeners grouped languages into the L-JPA classifications, counter to rhythm typology predictions. Thus, we still don't know the degree to which the L-JPA classifications are shared by different listeners.

3.1.2 Additional problems posed by rhythm metric approaches

3.1.2.1 Lack of explanatory power

Surface rhythm metric differences don't provide an explanation for why surface measure differences among languages exist. Although most of the rhythm metrics in the literature are thought to reflect rhythm-related structural attributes of different languages, it is often the case that any one given surface metric can be ambiguous as to the cause of the value it provides for any particular language, given that many different factors can affect the duration of any measured interval (Lindblom 1968; Lehiste 1970; Nooteboom 1972; Klatt 1976; van Santen 1992). For example, as discussed earlier in this paper and in Arvaniti (2009), the same value on the standard deviation of vocalic interval duration could arise from different sources, e.g., from vowel reduction, or from lengthening related to prosodic structure. Analyses which provide deeper insights into the reasons for rhythm metric differences among languages, e.g., by relating these differences to grammatical facts about each language, would be highly desirable. A related problem arises for using rhythm measures in areas where they have been proposed to be particularly useful, i.e., studies of disordered speech or of 2nd language productions. In these cases, given enough data, rhythm metrics might be used to identify a problem in the productions of certain individual speakers, but a definitive diagnosis of what the problem actually is would require additional studies designed to test specific hypotheses about the precise nature and underlying cause(s) of the timing anomalies. Given that additional studies would be required in order to diagnose the problem(s), the value of a rhythm metric study in such cases may be largely as a pointer to the need for additional analysis of which of the factors that influence speech timing are functioning inappropriately.

3.1.2.2 Rhythm metrics won't determine which of the definitions in Section 2 is most appropriate for speech

Even in conjunction with quantified impressions of rhythmic differences, rhythm metrics do not provide a definition of what speech rhythm is. If rhythm metrics correlate with quantified impressions of rhythmic differences, results of these studies would shed some light on what definition of this term is most appropriate for determining global rhythmic (or timing) profiles of different languages, but will not necessarily unequivocally provide one. For example, if syllable- vs. stress-timed languages are distinguished by position on a %V (proportion of speech taken up by vocalic intervals) continuum, a periodic definition might be ruled out, but it would be difficult to determine whether the pattern of %V was derivative of temporal computations of timing that refer to %V, or instead emerges from

timing computations in a different domain altogether. For example, a high %V value might result when a language uses a large number of [a] vowels, which are intrinsically long in duration due to the longer distance traveled by the articulators. A more productive approach to discovering an appropriate definition for speech rhythm might be to evaluate each of the definitions presented in Section 2 in turn. This would involve coming to understand how speech timing works, i.e., why speech timing is the way it is.

4 Why is speech timing the way it is?

Because the definitions in Section 2 all relate to speech timing, but in different ways, an important step toward understanding ways in which speech might be rhythmic, as well as how rhythmic characteristics might differ across languages, is to understand in broader terms how speech timing works. Understanding speech timing involves understanding the possible role of periodicity in motor control and perception (periodicity-based rhythm definitions 1b and 1c), the phonological structures that influence it (abstract phonological rhythm definition 2), as well as surface timing patterns (rhythm-as-surface-timing-patterns definition 3). While timing is not the only parameter of the speech signal that contributes to impressions of rhythm, as Barry, Andreeva, and Koreman (2009) and Arvaniti and Rodriquez (this volume) point out, it is a critical one, since impressions of rhythm are driven by ways in which parameters pattern in (abstract or surface) time.

We turn now to a discussion of what such a research program might look like. This approach involves addressing two broad questions:

1. What is the role of periodicity in speech timing control and perceptual processing? (discussed extensively above)
2. What types of cross-linguistic differences are there in the representations and processes that affect speech timing, and in the affected timing intervals?

Addressing this second question involves cross-linguistic phonological analysis, as well as careful empirical study of the ways in which phonological structures and categories affect the timing of different types of intervals, as well as other phonetic parameters. The cross-linguistic work that we advocate follows a large body of research on timing and the factors that influence it (see Fletcher 2010 for a review), which has not been consistently associated with the term ‘rhythm’. This research fits with the rhythm-as-surface-timing patterns definition (definition 3 above). Taken as a whole, it shows that controlled experiments and/or modeling studies with large labeled corpora (e.g., van Santen 1992) will be useful

for these cross-language comparisons, because disentangling which factor has caused or is largely responsible for a given surface phonetic (timing) result is difficult without controlling for confounding factors, which can be numerous.

Extensive work on languages like English, Dutch, Finnish, French, German, and Swedish has shown that a wide range of grammatical and performance factors influence speech timing. This range of factors can be usefully thought of in terms of two sets, one of which may determine the *global timing profile* of a language or language variety as a whole, while the other is more relevant in determining *the timing pattern for a particular utterance* in that language. The first set may include factors characteristic of the phonological and lexical structure of the language, i.e., (i) the inventory of contrastive phonological categories (e.g., phonemes and tonemes), including tense vs. lax vowels, phonemically short vs. long vowels and consonants, (ii) the relative frequency of each of these categories, (iii) the phonotactic structure of the language, i.e., the number and type of consonants that can appear in a sequence, given the syllable- and morpheme-structure constraints, and the relative frequency of each type of pattern, and (iv) word-prosody patterns, e.g., lexical stress and accent. This set also includes (v) language-specific phonetic correlates of phonological categories, e.g., correlates of lexical stress that differ cross-linguistically, (vi) language (variety)-specific ‘performance principles’ for implementing prosodic structure, e.g., differences in the magnitude of durational effects associated with prosodic structure (Cambier-Langeveld and Turk 1999; White, Payne, and Mattys 2009; Nakai et al. 2009; Nakai et al. 2012), and (vii) possibly, language-specific speech rate or stylistic factors (Krivokapić, this volume). The second set, which determines the timing characteristics of a given utterance of the language, includes factors which are utterance- and perhaps even speaker-specific. This set specifies (viii) the particular phonological units in the words of the utterance, (ix) their sequence, and (x) their prosodic structure (which itself may reflect the relative predictability of elements in the utterance; Aylett and Turk 2004; Turk 2010), as well as (xi) the speaker’s chosen speaking rate/style, and (xii) the degree of familiarity/practice the speaker may have with the words s/he is producing.

While there is good evidence for the influence of each of these factors on speech timing, less is known about the specifics of these structures and their relationships with other physical parameters, particularly across a range of languages. That is, which prominence and grouping structures are involved in individual languages, what are the magnitudes of the effects of these structures on timing and other rhythm-related parameters, and which stretches of speech do these structures affect? Preliminary answers to these questions can be derived from a comparison of single-language studies; however, because timing is affected by so many factors, apparent cross-linguistic differences found in such

studies can sometimes be spurious especially if the magnitudes of observed differences are subtle, and if different methods and/or materials have been used (see discussion in Cambier-Langeveld and Turk 1999). However, a growing body of evidence suggests that cross-linguistic differences in timing do exist (see Fletcher 2010 for a review). For example, Delattre (1966), Oller (1979), Dauer (1983), and Hoequist (1983), all cited in Fletcher 2010, suggest that prominence-related duration effects are smaller in Spanish than in English. Similarly, White, Payne, and Mattys (2009) showed that different varieties of Italian have different magnitudes of prosody-related lengthening. Subtle language-specific differences have also been found for the degree and distribution of accentual lengthening in Dutch vs. English (Cambier-Langeveld and Turk 1999). As more findings come to light, it will be important to determine the principles that govern language-specific timing effects. That is, are the differences qualitative, or quantitative? Can they be explained in terms of other aspects of the language systems that are involved, or are they arbitrary? For example, final lengthening in Finnish appears to be restricted on phonologically short vowels, presumably in order to preserve the quantity distinction (Nakai et al. 2009; Nakai et al. 2012). Similarly, polysyllabic shortening appears to be absent in Finnish and Estonian (Lehiste 1970; Suomi 2007), possibly because word boundary locations are signaled by fixed word-initial stress in these languages, and so don't require an additional durational cue. These observations suggest that timing adjustment phenomena may be related to factors such as the phonemic inventory and constraints on the possibilities for manipulating prominence and boundary-signaling cues of the language.

To sum up, while considerable research has been devoted to the search for diagnostic tools to capture the rhythmic differences that are widely believed to characterize different languages, the reliability and generality of these intuitions has not been tested. Moreover, it is possible that these differences will emerge more clearly from a study of the factors that influence speech timing than from a search for acoustic correlates of rhythm classes. We have suggested that these factors are large in number, and may operate differently in different languages, so that their investigation may provide a rich set of data on the sources of widely-held beliefs about rhythmic differences. The three papers in this section provide a taste of how such an approach might unfold. For example, as noted earlier, Arvaniti and Rodriquez (this volume) show that the *sasasa* approach to evaluating typological rhythm class membership does not produce the expected results, and suggest that a new approach might be useful. In this regard, they show that listeners are sensitive to differences in speaking rate, as well as surface timing and F_0 patterns relating to prosodic boundaries and prominences. Krivokapić (this volume) also shows that listeners are sensitive to rate differences,

as evidenced by their convergence behavior when speaking together. In doing so, she introduced the concept of what we are calling a timing profile at the utterance level, to which listeners are sensitive. We suggest it is also useful at the language-wide level, for characterizing the *global timing profile* of an entire language. Goswami and Leong (this volume) emphasize the importance of considering the role of temporal processing in perception and draw connections between this processing system and phonological processing, particularly during development. Thus, these papers illustrate the importance of considering new approaches to questions about speech rhythm. In this vein, we believe that a thorough consideration of the various factors that influence speech timing, and how they operate in different languages, will shed light on the question of how and whether languages differ systematically in characteristics that might be described as ‘rhythmic’. At the same time, we suggest that the use of the term ‘rhythm’ can be ambiguous, and suggest that the use of more explicit terms such as ‘periodicity’, ‘prosodic structure’, or ‘surface speech timing patterns’ might make research findings in this field easier to grasp.

References

- Abercrombie, David. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
- Ahissar, Ehud, Srikantan Nagarajan, Merav Ahissar, Athanassios Protopapas, Henry Mahncke, & Michael M. Merzenich. 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences* 98. 13367–13372.
- Arvaniti, Amalia. 2009. Rhythm, timing and the timing of rhythm. *Phonetica* 66. 46–63.
- Arvaniti, Amalia. 2012a. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40(3). 351–373.
- Arvaniti, Amalia. 2012b. Rhythm classes and speech perception. In O. Niebuhr (ed.), *Understanding Prosody: The Role of Context, Function and Communication*, 75–92. Berlin/Boston: De Gruyter.
- Arvaniti, Amalia, & Tara Rodriquez. 2013. The role of rhythm class, speaking rate, and F_0 in language discrimination. *Laboratory Phonology* 4(1). 7–38.
- Aylett, Matthew, & Alice Turk. 2004. The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration on spontaneous speech. *Language and Speech* 47. 31–56.
- Barbosa, Plinio A. 2007. From syntax to acoustic duration: A dynamical model of speech rhythm production. *Speech Communication* 49(9). 725–742.
- Barry, William J., Bistra Andreeva, & Jacques Koreman. 2009. Do rhythm measures reflect perceived rhythm? *Phonetica* 66. 78–94.
- Barry, William J., Bistra Andreeva, Michella Russo, Snezhina Dimitrova, & Tanja Kostadinova. 2003. Do rhythm measures tell us anything about language type? *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona.

- Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory, & Daniel Gildea. D. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113(2). 1001–1024.
- Bloch, Bernard. 1950. Studies in colloquial Japanese IV: phonemics. *Language* 26. 86–125.
- Bybee, Joan L., & Paul Hopper. 2001. *Frequency and the Emergence of Language Structure*. Amsterdam: John Benjamins.
- Byrd, Dani, & Elliot Saltzman. 2003. The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31(2). 149–180.
- Cambier-Langeveld, Tina, & Alice E. Turk. 1999. A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics* 27. 255–280.
- Casini, Laurence, Boris Burle, & Noël Nguyen. 2009. Speech perception engages a general timer: Evidence from a divided attention word identification task. *Cognition* 112(2). 318–322.
- Classe, André. 1939. *The Rhythm of English Prose*. Oxford: Basil Blackwell.
- Cummins, Fred. 2012. Oscillators and syllables: a cautionary note. *Frontiers in Psychology* 3. 364.
- Cutler, Anne. 1994. Segmentation problems, rhythmic solutions. *Lingua* 92(1–4). 81–104.
- Dasher, Richard, & Dwight Bolinger. 1982. On pre-accentual lengthening. *Journal of the International Phonetic Association* 12. 58–71.
- Dauer, R. M. 1983. Stress-timing and syllable-timing re-analysed. *Journal of Phonetics* 11. 51–62.
- Darwin, C. J., & Andrew Donovan. 1980. Perceptual studies of speech rhythm: isochrony and intonation. In J. C. Simon (ed.), *Proceedings of the NATO ASI on Spoken Language Generation and Understanding*, 77–95. Dordrecht: Reidel.
- Delattre, Pierre C. 1966. A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics* 4. 183–198.
- Dellwo, Volker. 2010. Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence. Ph.D. dissertation, Universität Bonn.
- Dellwo, Volker, & Petra Wagner. 2003. Relations between language rhythm and speech rate. *Proceedings of the 15th International Congress of Phonetic Sciences*, 471–474. Barcelona.
- den Os, Elisabeth Anna. 1988. *Rhythm and Tempo of Dutch and Italian: A Contrastive Study*. University of Utrecht.
- Deterding, David. 2001. The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics* 29. 217–230.
- Donovan, Andrew, & C. J. Darwin. 1979. The perceived rhythm of speech. *Proceedings of the Ninth International Congress of Phonetic Sciences*, Vol. 2, 268–274. Copenhagen.
- Fant, Gunnar, & A. Kruckenberg. 1989. Preliminaries to the study of Swedish prose reading and reading style. *STL-QPSR* 2. 1–83.
- Fletcher, Jane. 2010. The prosody of speech: timing and rhythm. In William J. Hardcastle, John Laver, & Fiona E. Gibbon (eds.), *The Handbook of Phonetic Sciences* (2nd ed.), 521–602. West Sussex, UK: Blackwell.
- Ghitza, Oded. 2011. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology* 2. Article 130.
- Giegerich, Heinz J. 1980. On stress-timing in English phonology. *Lingua* 51, 187–221.

- Giraud, Anne-Lise, & David Poeppel. 2012. Speech perception from a neurophysiological perspective. In David Poeppel, Tobias Overath, Arthur Popper, & Richard Fay (eds.), *Human Auditory Cortex*. New York: Springer.
- Goswami, Usha, & Victoria Leong. 2013. Speech rhythm and temporal structure: Converging perspectives? *Laboratory Phonology* 4(1). 67–92.
- Grabe, Esther, & Ee Ling Low. 2002. Durational variability in speech and the Rhythm Class Hypothesis. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology* 7. Berlin: Mouton de Gruyter.
- Han, Mieko S. 1962. The feature of duration in Japanese. *Onsei no kenkyuu (Study Sounds)* 10. 65–80.
- Hoequist, Charles, Jr. 1983. Syllable duration in stress-, syllable- and mora-timed languages. *Phonetica* 40. 203–237.
- Jurafsky, Daniel, Alan Bell, Michelle Gregory, & William Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In Joan Bybee & Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, 229–254. Amsterdam: John Benjamins.
- Kim, Heejin. 2006. Speech rhythm in American English: A corpus study. Ph.D. dissertation, University of Illinois at Urbana-Champaign.
- Kim, Heejin, & Jennifer Cole. 2005. The stress foot as a unit of planned timing: evidence from shortening in the prosodic phrase. *Proceedings of Interspeech 2005*, Lisbon, Portugal.
- Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59(5). 1208–1221.
- Kohler, Klaus. 2009. Rhythm in speech and language: a new research paradigm. *Phonetica* 66. 29–45.
- Kozhevnikov, Valeriy A., & Ludmilla A. Chistovich. 1965. Speech: articulation and perception. Joint Publ. Res. Serv., Washington, D.C.
- Krivokapić, Jelena. 2013. Rhythm and convergence between speakers of American and Indian English. *Laboratory Phonology* 4(1). 39–65.
- Ladefoged, Peter. 1975. *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.
- Large, Edward W. 2010. Neurodynamics of music. In Mari Riess Jones, Richard R. Fay, & Arthur N. Popper (eds.), *Springer Handbook of Auditory Research*, Vol. 36: Music Perception, 201–231. New York: Springer.
- Lea, Wayne A. 1974. Prosodic aids to speech recognition IV. A general strategy for prosodically-guided speech understanding *Univac Report No. PX10791*. St. Paul, Minnesota: Sperry Univac, DSD.
- Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.
- Lehiste, Ilse. 1972. The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America* 51(6, part 2). 2018–2024.
- Lehiste, Ilse. 1977. Isochrony reconsidered. *Journal of Phonetics* 5. 253–264.
- Lieberman, Mark. 1975. The intonational system of English. Ph.D. dissertation, MIT.
- Lieberman, Mark, & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249–336.
- Lindblom, Björn. 1968. Temporal organization of syllable production. *Reports of the 6th International Congress of Acoustics*, B29–B30. Tokyo.
- Lloyd James, Arthur. 1935. *The Broadcast Word*. London: Kegan Paul, Trench, Trubner & Co.
- Lloyd James, Arthur. 1940. *Speech Signals in Telephony*. London: Pitman.

- Loehr, Janeen, Edward W. Large, & Caroline Palmer. 2011. Temporal coordination in music performance: Adaptation to tempo change. *Journal of Experimental Psychology: Human Perception and Performance* 37(4). 1292–1309.
- Loukina, Anastassia, Greg Kochanski, Burton Rosner, Elinor Keane, & Chilin Shih. 2011. Rhythm measures and dimensions of durational variation in speech. *Journal of the Acoustical Society of America* 129(5). 3258–3270.
- Low, Ee Ling, Esther Grabe, & Francis Nolan. 2000. Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech* 43. 377–402.
- Mehler, Jacques, Emmanuel Dupoux, Thierry Nazzi, & Ghislaine Dehaene-Lambertz. 1996. Coping with linguistic diversity: the infant's viewpoint. In James L. Morgan & Katherine Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, 101–116. Mahwah, NJ: Lawrence Erlbaum Associates.
- Miller, Joanne L. 1981. Effects of speaking rate on segmental distinctions. In Peter Eimas & Joan L. Miller (eds.), *Perspectives in the Study of Speech*. New Jersey: Erlbaum Press.
- Murty, Lalita, Takashi Otake, & Anne Cutler. 2007. Perceptual tests of rhythmic similarity: I. Mora rhythm. *Language and Speech* 50(1). 77–99.
- Nakai, Satsuki, Sari Kunnari, Alice Turk, Kari Suomi, & Riikka Ylitalo. 2009. Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics* 37. 29–45.
- Nakai, Satsuki, Alice Turk, Kari Suomi, Sonia Granlund, Riikka Ylitalo, & Sari Kunnari. 2012. Quantity constraints on the temporal implementation of phrasal prosody in Northern Finnish. *Journal of Phonetics* 40. 796–807.
- Nespor, Marina. 1990. On the rhythm parameter in phonology. In I. M. Roca (ed.), *Logical Issues in Language Acquisition*, 157–175. Dordrecht: Foris.
- Newman, Rochelle, & James R. Sawusch. 1996. Perceptual normalization for speaking rate: Effects of temporal distance. *Perception and Psychophysics* 58(4). 540–560.
- Nolan, Francis, & Eva Liina Asu. 2009. The Pairwise Variability Index and coexisting rhythms in language. *Phonetica* 66. 64–77.
- Nolan, Francis, & Hae-Sung Jeon. To appear. Speech rhythm: a mirage? Paper presented at Perspectives on Rhythm and Timing, Glasgow.
- Nooteboom, Sibout Govert. 1972. Production and perception of vowel duration: A study of durational properties in Dutch. Ph.D. dissertation, University of Utrecht.
- Nourski, Kirill V., Richard A. Reale, Hiroyuki Oya, Hiroto Kawasaki, Christopher K. Kovach, Haiming Chen, Matthew A. Howard, III, & John F. Brugge. 2009. Temporal envelope of time-compressed speech represented in the human auditory cortex. *Journal of Neuroscience* 29, 15564–15574.
- O'Connor, Joseph D. 1968. The duration of the foot in relation to the number of component sound segments. *Progress Report 3*. Phonetics Laboratory, University College, London.
- O'Dell, Michael L., & Tommi Nieminen. 1999. Coupled oscillator model of speech rhythm. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville, & Ashlee C. Bailey (eds.), *Proceedings of the XIVth International Congress of Phonetic Sciences*, Vol. 2, 1075–1078. New York: American Institute of Physics.
- Oller, D. Kimbrough. 1979. Syllable timing in Spanish, English, and Finnish. In Harry Hollien & Patricia Hollien (eds.), *Current Issues in the Phonetic Sciences*, 320–341. Amsterdam: John Benjamins.
- Patel, Aniruddh D. 2008. *Music, Language and the Brain*. Oxford University Press.
- Pike, Kenneth L. 1945. *The Intonation of American English*. Ann Arbor: The University of Michigan Press.

- Ramus, Franck. 2002. Acoustic correlates of linguistic rhythm: Perspectives. Paper presented at the Proceedings of Speech Prosody 2002, Aix-en-Provence.
- Ramus, Franck, Emmanuel Dupoux, & Jacques Mehler. 2003. The psychological reality of rhythm classes: Perceptual studies. Paper presented at the Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona.
- Ramus, Franck, & Jacques Mehler. 1999. Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America* 105. 512–521.
- Ramus, Franck, Marina Nespor, & Jacques Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 72. 1–28.
- Reinisch, Eva, Alexandra Jesse, & James McQueen. 2011. Speaking rate affects the perception of duration as a suprasegmental lexical-stress cue. *Language and Speech* 54(2). 147–165.
- Renwick, Margaret E. L. 2012. What does %V actually measure? *Cornell Working Papers in Phonetics and Phonology* 2012, 1–20.
- Roach, Peter. 1982. On the distinction between “stress-timed” and “syllable-timed” languages. In David Crystal (ed.), *Linguistic Controversies*. London: Edward Arnold.
- Saltzman, Elliot, Hosung Nam, Jelena Krivokapić, & Louis Goldstein. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. Paper presented at Speech Prosody 2008, Campinas, Brazil.
- Salverda, Anne Pier, Delphine Dahan, & James McQueen. 2003. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90. 51–89.
- Scott, Donia R., S. D. Isard, & Bénédicte de Boysson-Bardies. 1985. Perceptual isochrony in English and in French. *Journal of Phonetics* 13. 155–162.
- Selkirk, Elisabeth O. 1984. *Phonology and Syntax: The Relation Between Sound and Structure*. Cambridge, MA: MIT Press.
- Shattuck-Hufnagel, Stefanie, & Alice Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25(2). 193–247.
- Shattuck-Hufnagel, Stefanie, & Alice Turk. 2011. Durational evidence for word-based vs. prominence-based constituent structure in limerick speech. *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, China.
- Shatzman, Keren, & James McQueen. 2006. Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics* 68(1). 1–16.
- Suomi, Kari. 2007. On the tonal and temporal domains of accent in Finnish. *Journal of Phonetics* 35. 40–55.
- Thompson, Henry Swift. 1980. Stress and salience in English: theory and practice. Ph.D. dissertation, University of California.
- Turk, Alice. 2010. Does prosodic constituency signal relative predictability? A Smooth Signal Redundancy hypothesis. *Journal of Laboratory Phonology* 1. 227–262.
- Turk, Alice. 2012. The temporal implementation of prosodic structure. In Abigail C. Cohn, Cécile Fougeron, & Marie K. Huffman (eds.), *Oxford Handbook of Laboratory Phonology*, 242–253.
- Turk, Alice E., & Stefanie Shattuck-Hufnagel. 2000. Word-boundary-related duration patterns in English. *Journal of Phonetics* 28. 397–440.
- Turk, Alice E., & Stefanie Shattuck-Hufnagel. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics* 35(4). 445–472.
- Turk, Alice E., & Laurence White. 1999. Structural influences on accentual lengthening in English. *Journal of Phonetics* 27. 171–206.

- Uldall, Elizabeth T. 1971. Isochronous stresses in R.P. In Louis Leonor Hammerich, Roman Jakobson, & Eberhard Zwirner (eds.), *Form and Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jorgensen*. Akademisk Forlag.
- van Santen, Jan P. H. 1992. Contextual effects on vowel duration. *Speech Communication* 11. 513–546.
- van Santen, Jan, & Chilin Shih. 2000. Suprasegmental and segmental timing models in Mandarin Chinese and American English. *Journal of the Acoustical Society of America* 107(2). 1012–1026.
- White, Laurence, & Sven L. Mattys. 2007. Calibrating rhythm: First language and second language studies. *Journal of Phonetics* 35. 501–522.
- White, Laurence, Sven L. Mattys, & Lukas Wiget. 2012. Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language* 66. 665–679.
- White, Laurence, Elinor Payne, & Sven L. Mattys. 2009. Rhythmic and prosodic contrast in Venetan and Sicilian Italian. In Marina Vigário, Sónia Frota, and M. João Freitas (eds.), *Phonetics and Phonology: Interactions and interrelations*, 137–158. Amsterdam: John Benjamins
- White, Laurence, & Alice Turk. 2010. English words on the Procrustean bed: polysyllabic shortening reconsidered. *Journal of Phonetics* 38. 459–471.
- Wiget, Lukas, Laurence White, L., Barbara Schuppler, Izabelle Grenon, Olesya Rauch, & Sven Mattys. 2010. How stable are acoustic metrics of contrastive speech rhythm? *Journal of the Acoustical Society of America* 127(3). 1559–1569.
- Wightman, Colin W., Stefanie Shattuck-Hufnagel, Marie Ostendorf, & Patti J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91(3). 1707–1717.
- Williams, Briony, & S. M. Hiller. 1994. The question of randomness in English foot timing: a control experiment. *Journal of Phonetics* 22. 423–439.
- Wing, Alan M., & Alfred B. Kristofferson. 1973. Response delays and the timing of discrete motor responses. *Perception & Psychophysics* 14(1). 5–12.